ENABLING NEXT GENERATION ETHERNET ACCESS WITH ETHERNET PASSIVE OPTICAL NETWORKS

Gerry Pesavento, President and CEO, Teknovus, gerry.pesavento@teknovus.com
Glen Kramer, System Architect, Teknovus, glen.kramer@teknovus.com

Abstract

This article describes Ethernet Passive Optical Networks (EPONs), an emerging local subscriber access architecture that combines low cost point-to-multipoint fiber infrastructure with Ethernet. EPONs are designed to carry Ethernet frames at standard Ethernet rates. EPONs use a single trunk fiber that extends from a central office to a passive optical splitter, which then fans out to multiple optical drop fibers connected to subscriber nodes. Other than the end terminating equipment, no component in the network requires electrical power, hence the term passive. Local carriers have long been interested in passive optical networks for the benefits they offer: minimal fiber infrastructure and no powering requirement in the outside plant. With Ethernet now emerging as the protocol of choice for carrying IP traffic in the Metro and Access network, EPON has emerged as a potentially optimal architecture for fiber to the building and fiber to the home.

1 Introduction

While in recent years the telecommunications backbone has experienced substantial growth, little has changed in the access network. The "last mile" still remains the bottleneck between high capacity Local Area Networks (LANs) and the backbone network.

The most widely deployed "broadband" solutions today are Digital Subscriber Line (DSL) and Cable Modem (CM) networks. Although they are an improvement compared to dial-up modems, they are unable to provide enough bandwidth for emerging services such as IP telephony, Video-On-Demand (VoD), interactive gaming or two-way video conferencing. Both DSL and Cable modem architectures are built on top of existing copper/coax communication infrastructures, which are not optimized for data traffic. In cable modem networks, only a few RF channels are dedicated for data, while the majority of bandwidth is tied up servicing legacy broadcast video. DSL copper networks do not allow sufficient data rates at required distances. Most network operators have come to the realization that a new technology is required: one that is inexpensive, simple, scalable, and capable of delivering bundled voice, data and video services to an end-user subscriber over a single data-centric network. Ethernet Passive Optical Networks (EPONs), which represent the convergence of ubiquitous Ethernet equipment and a low-cost fiber infrastructure, appear to be the best candidate for the next generation access network.

2 Next Generation Access Network

Optical fiber is capable of delivering bandwidth intensive, integrated voice, data and video services at distances beyond 20 kilometers in the subscriber access network. A logical way to deploy optical fiber in the local access network is using a Point-to-Point (P2P) topology, with dedicated fiber runs from the Local Exchange to each end-user subscriber (Figure 1.a). While this is a simple architecture, in most cases it is cost prohibitive because it requires significant outside plant fiber deployment as well as connector termination space in the Local Exchange. Considering N subscribers at an average distance L km from the central office, a P2P design requires $2N$ transceivers and $N*L$ total fiber length (assuming a single fiber is used for bi-directional transmission).

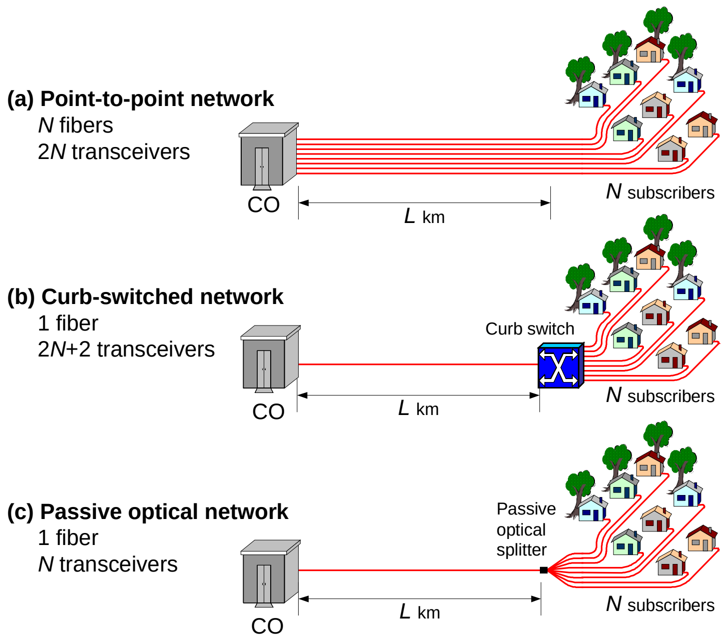


Figure 1: Fiber to the home (FTTH) deployment scenarios

To reduce fiber deployment, it is possible to deploy a remote switch (concentrator) close to the neighborhood. This reduces the fiber consumption to only L km (assuming negligible distance between the switch and customers), but actually increases the number of transceivers to $2N+2$, as there is one more link added to the network (Figure 1.b). In addition, a curb-switched architecture requires electrical power as well as back-up power at the curb unit. Currently, one of the highest costs for local exchange carriers is providing and maintaining electrical power in the local loop.

Therefore, it is logical to replace the hardened active curb-side switch with an inexpensive passive optical component. Local carriers have long been interested in PONs for the minimal number of optical transceivers, central office terminations and fiber deployment they require compared to P2P solutions. A PON is a point-to-multipoint optical network with no active elements in the signal's path from source to destination. The only interior elements used in a PON are passive optical components, such as optical fiber, splices and splitters. Access networks based on a single fiber PON only require $N + 1$ transceivers and L km of fiber (Figure 1.c).

3 EPON Architecture

Ethernet PON (EPON) is a passive optical network that carries data traffic encapsulated in Ethernet frames as defined in the IEEE 802.3 standard draft [1]. EPON uses a standard 8b/10b line coding and operates with standard Ethernet frames and at standard Ethernet speed of 1 Gbps.

All transmissions in a PON are performed between the Optical Line Terminal (OLT) and Optical Network Units (ONU). The OLT resides in the local exchange (central office), connecting the optical access network to the metro backbone. The ONU is located either at the curb (FTTC solution), or at the end-user location (FTTH and FTTB), and provides broadband voice, data, and video services. In the downstream direction (from OLT to ONUs), a PON is a point-to-multipoint network, and in the upstream direction it is a multipoint-to-point network.

3.1 Principles of Operation

In the downstream direction (from the OLT to ONUs), Ethernet packets transmitted by the OLT pass through a $1{\times}N$ passive splitter and reach each ONU. The value of N is typically between 4 and 64 (limited by the available power budget). Each ONU extracts only the packets that are designated for it or the subscribers attached to it, and discards the rest of them (see Figure 2).

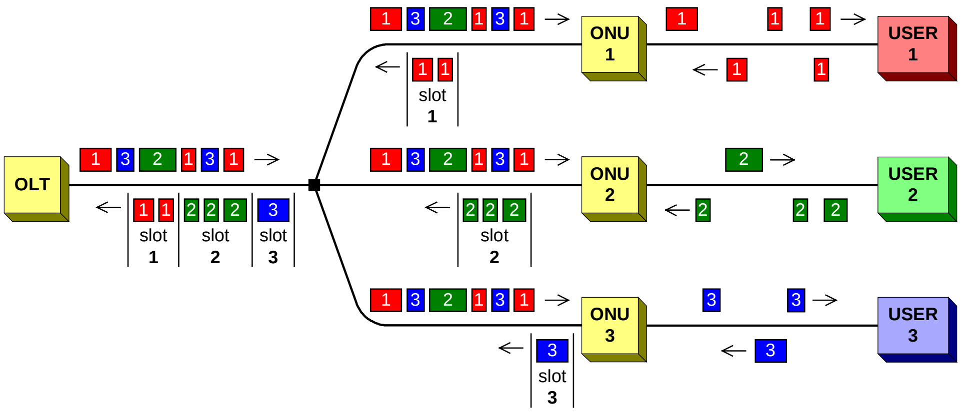


Figure 2: Downstream transmission in an EPON

In the upstream direction (from ONUs to the OLT), each ONU is allocated a timeslot. Each timeslot is capable of carrying several Ethernet packets. An ONU should buffer frames received from a subscriber until its timeslot arrives. When its timeslot arrives, the ONU will "burst" all stored frames at the full channel speed of 1 Gbps. If there are no frames in the buffer to fill the entire timeslot, 10-bit idle characters are transmitted (as specified for full-duplex Ethernet).

The performance of an EPON depends on the particular slot-allocation scheme. The possible timeslot allocation schemes could range from static allocation, such as fixed time-division multiple access (TDMA), to schemes dynamically adapting the slot size based on instantaneous queue load in every ONU (statistical multiplexing schemes).

3.2 Multi-Point Control Protocol (MPCP)

To support a timeslot allocation by the OLT, the multi-point control protocol (MPCP) is being developed by the 802.3ah task force. MPCP is not concerned with a particular bandwidth-allocation (or inter-ONU scheduling) scheme; rather it is a supporting mechanism that can facilitate implementation of various bandwidth-allocation algorithms in EPON.

This protocol relies on two Ethernet messages: GATE and REPORT[1]. A GATE message is sent from the OLT to an ONU and it used to assign a transmission timeslot. A REPORT message is used by an ONU to convey its local conditions (such as buffer occupancy, etc.) to the OLT and to help the OLT make intelligent allocation decisions. Both GATE and REPORT messages are MAC control frames (type 88-08) and are processed by the MAC control sub-layer.

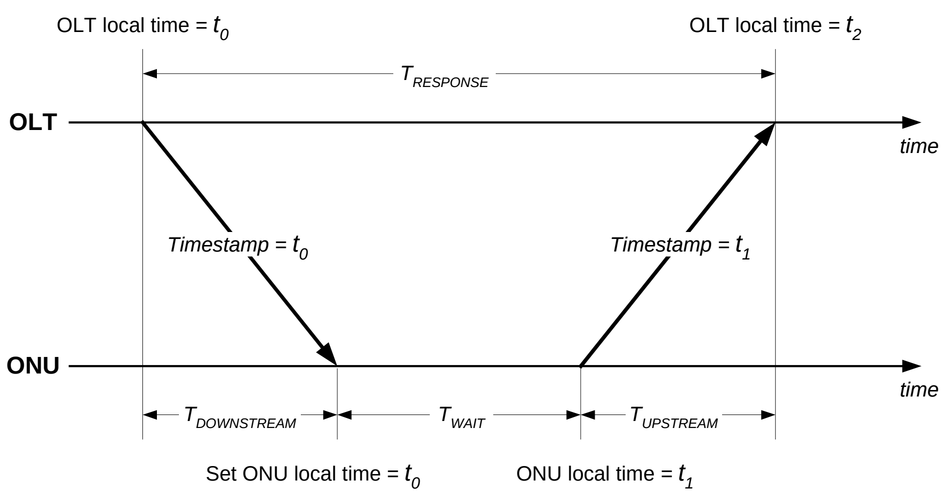
[1] Additionally MPCP defines REGISTER_REQUEST, REGISTER, and REGISTER_ACK messages used for ONU's registration. For more information, please refer to [2].

Below, we illustrate the normal operation of MPCP:

1. From its higher layer (MAC control client), MPCP in the OLT gets a request to transmit a GATE message to a particular ONU with the following information: time when that ONU should start transmission and length of the transmission.
2. The MPCP layer (in the OLT and each ONU) maintains a clock. Upon passing a GATE message from its higher layer to the MAC, MPCP timestamps the message with its local time.
3. When an ONU receives a GATE message matching its MAC address (GATE messages are unicast), it programs its local registers with transmission start and transmission length values obtained from the GATE message. The ONU will also update its local clock to that of the timestamp extracted from the received GATE message.
4. When the local time reaches the '*start transmission*' register value, the ONU will start transmitting. That transmission may include multiple Ethernet frames. The ONU will ensure that no frames are fragmented. If the next frame does not fit in the remainder of the timeslot, it will be deferred till the next timeslot, leaving some unused remainder in the current timeslot.

REPORT messages are sent by ONUs in the assigned transmission windows together with data frames. REPORT messages can be sent automatically or on the OLT's demand. A REPORT message is generated in the MAC control client layer and is time stamped in the MAC control. Typically, the REPORT message would contain the desired size of the next timeslot based on the ONU's queue size. When requesting a timeslot, an ONU should account for additional overhead, namely the 64-bit frame preamble and the 96-bit inter-frame gap (IFG) associated with every Ethernet packet.

When a time stamped REPORT message arrives at the OLT, it is passed to the MAC control client layer responsible for making the bandwidth-allocation decision. Additionally, the OLT will recalculate the round-trip time (RTT) to the source ONU. As shown in Figure 3, the RTT equals exactly the difference between the REPORT arrival time and the timestamp contained in the REPORT message.



$T_{DOWNSTREAM}$ = downstream propagation delay

$T_{UPSTREAM}$ = upstream propagation delay

T_{WAIT} = wait time at ONU ($T_{WAIT} = t_1 - t_0$)

$T_{RESPONSE}$ = response time at OLT ($T_{RESPONSE} = t_2 - t_0$)

$$RTT = T_{DOWNSTREAM} + T_{UPSTREAM} = T_{RESPONSE} - T_{WAIT} = (t_2 - t_0) - (t_1 - t_0) = t_2 - t_1$$

Figure 3: Round-trip time measurement

Some small deviation of the new RTT from the previously measured RTT may be caused by changes in the fiber refractive index resulting from temperature drift. A large deviation should alarm the OLT about the ONU's potential mis-synchronization and should prevent the OLT from further granting any transmissions to that ONU until it is reinitialized (resynchronized).

The above description represents a framework of the protocol being developed for the EPON. For a detailed overview of MPCP operation, please consult [1].

3.3 Topology Emulation Sub-layer

To ensure compatibility with the 802.1D standard and guarantee seamless integration with other Ethernet networks, devices attached to the EPON medium will have an additional sub-layer that emulates a point-to-point medium. This sub-layer resides below the MAC sub-layer to preserve the existing Ethernet MAC operation defined in the IEEE standard P802.3. Operation of the emulation layer relies on tagging of Ethernet frames with tags unique to each ONU (Figure 4). These tags are called "link ID" and are placed in the preamble before each frame. To guarantee the uniqueness of link IDs, each ONU is assigned one or more tags by the OLT during the initial registration phase.

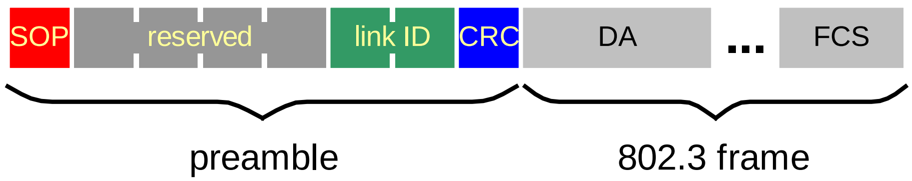
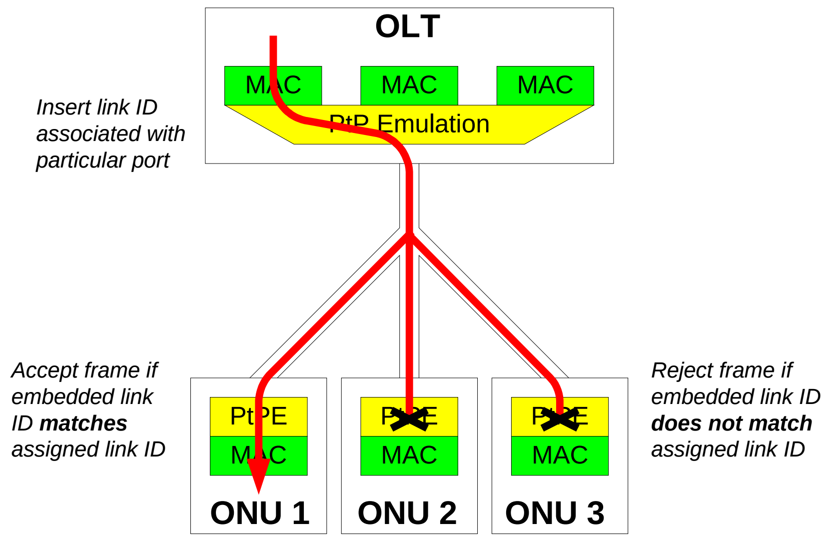


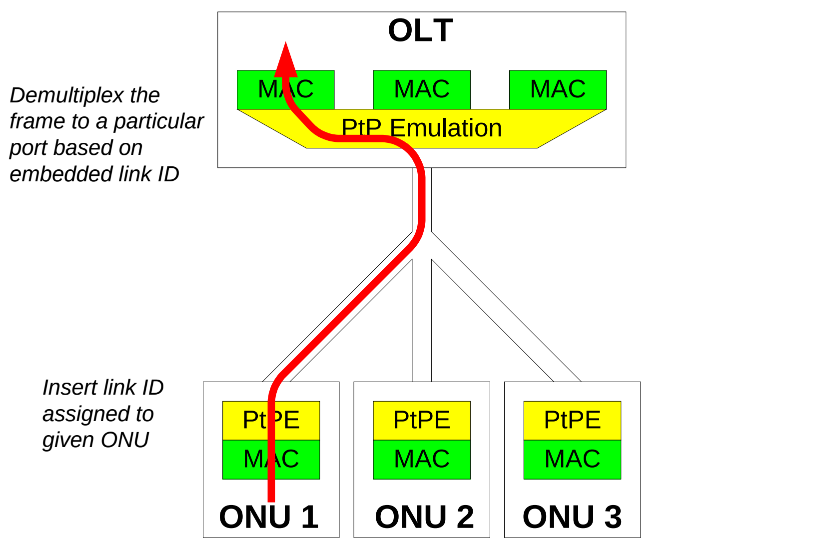
Figure 4: Link ID field embedded in frame preamble

To emulate a point-to-point medium, the OLT must have *N* MAC ports (interfaces), one for each ONU (Figure 5). When sending a frame downstream, the emulation sub-layer in the OLT will insert the link ID associated with a particular MAC port that the frame arrived from (Figure 5.a). Even though the frame will be delivered to each ONU, only one ONU will match that frame's link ID with the value assigned to the ONU and will accept the frame and pass it to its MAC layer for further verification. MAC layers in all other ONUs will never see that frame. In this sense, it appears as if the frame was sent on a point-to-point link to only one ONU.

In the upstream direction, the ONU will insert its assigned link ID in the preamble of each transmitted frame. The emulation sub-layer in the OLT will de-multiplex the frame to the proper MAC port based on the unique link ID (Figure 5.b).



(a) Downstream Transmission



(b) Upstream Transmission

Figure 5: Point-to-point emulation

4 EPON Efficiency

EPON efficiency depends on many parameters, such as packet size distribution, configuration of the scheduler, and the speed of the laser driver and clock recovery circuits, etc. Making unrealistic assumptions about any of these parameters can result in efficiency numbers being far off from the true value. It is, therefore, clear that to answer the question of EPON efficiency, one has to come up with an unambiguous set of EPON operational parameters and traffic characteristics. In this section, we identify all the parameters affecting the efficiency and justify the chosen values for these parameters.

By network efficiency we usually mean the *throughput efficiency*, also called *utilization*. Throughput is a measure of how much user data (application-level data) the network can carry through in a unit of time. Throughput efficiency is a ratio of the maximum throughput to the network bit rate. Perhaps the easiest way to calculate the efficiency is to find the overhead components associated with encapsulation and scheduling.

4.1 Encapsulation Overhead

The Ethernet encapsulation (framing) overhead is a result of adding the 8-byte frame preamble, 14-byte Ethernet header, and 4-byte FCS field to the MAC Service data units comprised of users' data. Additionally, at least a 12-byte minimum inter-frame gap (IFG) should be left between two adjacent frames[2]. Thus, the absolute overhead per one frame is constant and equal to 38 bytes[3] (see Figure 6). This encapsulation overhead is not specific to EPON, but a property common to all Ethernet networks.

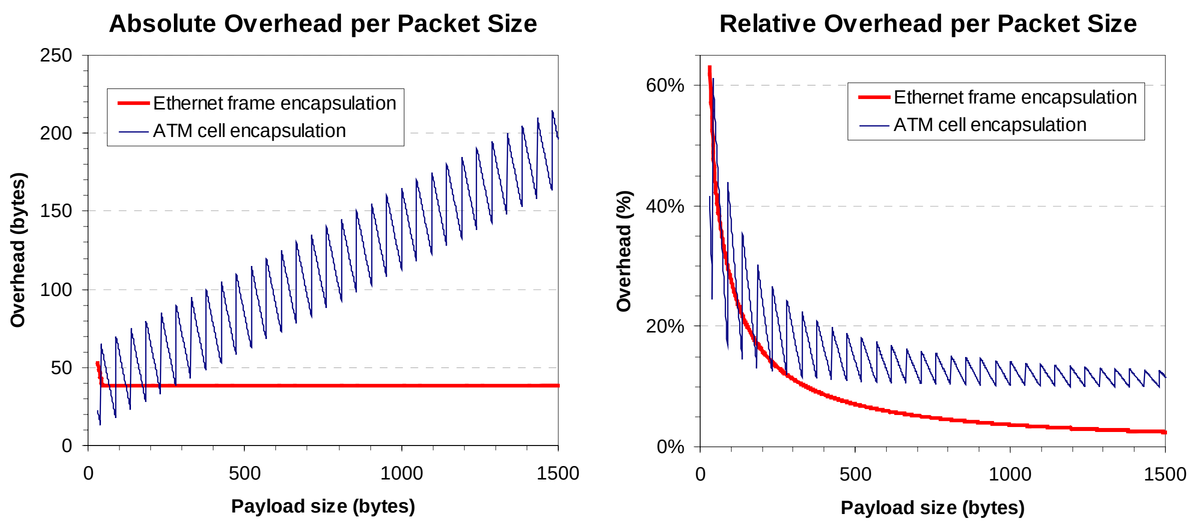


Figure 6: Comparison of Ethernet framing overhead and ATM cell tax

In ATM networks, the user's data units, such as IP datagrams, should be broken into multiple cells. The ATM encapsulation overhead (also known as *cell tax*) is comprised of multiple cell headers, an 8-byte ATM Adaptation Layer 5 (AAL5) trailer, and variable-size padding. The AAL5 trailer is needed for proper IP-datagram reassembly, and the padding is used to fill any remaining portion of the last cell. As is seen in Figure 6, the ATM encapsulation overhead depends on the payload size and is considerably higher than the Ethernet overhead.

The average value of the encapsulation overhead depends on the distribution of packet sizes. The distributions of packet sizes were reported in the literature. These distributions have a tri-modal shape and are similar for backbone networks [3] and access networks [4].

With the packet-size distribution obtained in a head-end of a cable network, the Ethernet encapsulation overhead is equal to 4.42%. Using the same distribution of packet sizes makes the average ATM encapsulation overhead equal to 13.22%. The encapsulation overhead shows the advantages of using variable-sized Ethernet frames to carry variable-sized IP packets. The Ethernet frame encapsulation overhead of 7.42% is lower than the ATM cell encapsulation overhead of 13.22%.

---

2 IFG is specified as 96-ns time interval, which is equal to 12 byte-transmission times in 1Gbps (1000BASE-X) Ethernet.

3 Short payloads are padded to a minimum length of 46-bytes. This also contributes to the Ethernet encapsulation overhead and is counted in our calculations.

4.2 Scheduling Overhead

The scheduling overhead in an EPON consists of control message overhead, guard band overhead, discovery overhead, and frame delineation overhead. In the downstream direction, only the control message overhead is present.

Control message overhead represents bandwidth lost due to the use of in-band control messages such as GATEs and REPORTs. The amount of overhead depends on the number of ONUs and the cycle time, i.e., an interval of time in which each ONU should receive a GATE message and send a REPORT message. We make an assumption here that the scheduling algorithm requires only one GATE message and one REPORT message to be exchanged between each ONU and the OLT in one cycle time.

Guard band overhead depends on PMD and PMA parameters such as Laser ON/OFF times, Automatic Gain Control (AGC) and Clock-and-Data Recovery (CDR) times. The draft IEEE 802.3ah D1.414 specifies four possible values (classes) for the AGC and CDR parameters: 96 ns, 192 ns, 288 ns, and 400 ns. The Laser ON/OFF times are fixed at 512 ns. In addition, guard bands should include a 128-ns dead zone to allow for timing variability of the Multi-Point Control Protocol[4]. As is shown in Figure 7, the Laser OFF time may partially overlap the laser ON time of the next ONU.

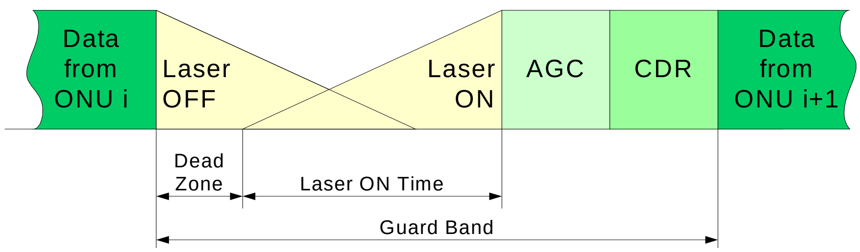


Figure 7: Structure of the guard band

Discovery overhead represents the bandwidth lost due to allocation of a discovery window. The discovery window should be larger than the maximum round-trip time of 200 μs. In our calculations we assume a discovery window size of 300 μs. The frequency of the discovery attempts is not specified in the IEEE 802.3ah draft. Intelligent algorithms may detect a situation when all ONUs are operational and cease all discovery attempts. We, however, will assume a simpler algorithm that performs periodic discovery regardless of the number of registered ONUs. The discovery period can be very large, for example 1 second or more. With a 1-second discovery period, the discovery overhead is equal 300 μs / 1 second = 0.03%.

Frame delineation overhead is associated with the fact that variable-sized frames may not be able to completely occupy the fixed-sized cycle. Grants to ONUs are based on their reported queue lengths. However, multiple grants with their associated guard bands may not fill the fixed cycle time exactly. The formula for the expected size of the unused remainder for an arbitrary packet size distribution was derived in [5].

Using the packet size distribution from [4], we get the average remainder approximately equal to 595 bytes. This means that we should expect, on average, 595 bytes wasted due to variable-sized frames not packing the fixed cycle completely.

Table 1 summarizes the values of various overhead components and calculates the combined upstream overhead. The presented values were calculated for a configuration with 32 ONUs and 1 ms cycle time.

[4] The presented timing values are based on IEEE802.3ah draft D1.414.

	Downstream transmission	Upstream transmission
Control message overhead	**2.15 %**	**2.15 %**
Guard band overhead	**0**	**2.66 %**
Discovery overhead	**0**	**0.03 %**
Frame delineation overhead	**0**	**0.48 %**
Total scheduling overhead	**2.15 %**	**5.32 %**

Table 1: Summary of scheduling overhead

Thus, EPON efficiency in the upstream direction is 94.68 % compared to a 1GbE point-to-point link. In the downstream direction, EPON efficiency reaches 97.85% of the efficiency of a point-to-point 1GbE link.

It is possible that a particular scheduling algorithm or implementation will have lower efficiency, however, that would only be the result of particular design decisions and not an intrinsic overhead of EPON architecture.

5 Conclusion

We have shown that EPON is a bandwidth efficient optical access network. EPON is ideally suited for IP data services such as VOIP, VPN, Internet access and IP Video. Applying appropriate dynamic bandwidth allocation schemes and quality of service mechanisms, EPON can also support time critical services such as POTS and TDM circuits, making EPON a full-service access network.

While the Multi-Point Control Protocol (MPCP) is well defined for EPON, there are many system level issues that are important for a successful optical access network implementation. These system level issues include common management system requirements, and interoperable security definitions and authentication techniques.

The promise of a low-cost first mile Ethernet access network supporting Ethernet over copper, point to point Ethernet and EPON is here. For the first time, network operators have a wide choice of interoperable broadband deployment solutions, which can be bound under a common management platform. EPON, because it allows for lower cost fiber infrastructure, will be an important part of the first mile.

References

[1] IEEE Standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks – Specific Requirements. Part 3: *Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specification*, ANSI/IEEE Std 802.3-2002, 2002 Edition. Available at *http://standards.ieee.org/getieee802/download/802.3-2002.pdf*.

[2] G. Kramer, B. Mukherjee, and A. Maislos, *Ethernet Passive Optical Networks*, In S. Dixit, editor, IP over WDM: Building the Next Generation Optical Internet, John Wiley & Sons, Inc., February 2003.

[3] K. Claffy, G. Miller, and K. Thompson, "*The nature of the beast: Recent traffic measurements from an internet backbone*," in Proceedings INET '98, (Geneva, Switzerland), July 1998.

[4] D. Sala and A. Gummalla, "PON Functional Requirements: Services and Performance," presented at IEEE 802.3ah meeting in Portland, OR, July 2001. Available at: *http://grouper.ieee.org/groups/802/3/efm/public/jul01/presentations/sala_1_0701.pdf.*

[5] G. Kramer, B. Mukherjee, and G. Pesavento, "*Ethernet PON (ePON): Design and Analysis of an Optical Access Network*," Photonic Network Communications, vol. 3, no. 3, pp. 307-319, July 2001.